DOCUMENT RESUME

ED 032 647 24 EA 002 553

By-Glaser, Robert

Evaluation of Instruction and Changing Educational Models.

California Univ., Los Angeles. Center for the Study of Evaluation.; Pittsburgh Univ., Pa. Learning Research and Development Center.

Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.

Report No-CSEIP-Occ-Rep-13; Reprint-46

Bureau No-BR-5-0253; BR-6-1646

Pub Date Sep 68

Contract-OEC-4-6-061646-1909; OEC-4-10-158

Note-28p.; From the Proceedings of the Symposium on Problems in the Evaluation of Instruction (Univ. of Calif., Los Angeles, December, 1967). Reprint.

EDRS Price MF-$0.25 HC-$1.50

Descriptors-Change Agents, Cognitive Development, Continuous Learning, *Educational Change, Educational Diagnosis, Educational Objectives, Educational Specifications, *Educational Theories, *Evaluation Needs, *Evaluation Techniques, Individualized Instruction, *Models

Changing educational practices require changes in our theories and techniques of evaluation. Three forces of change are: (1) The emphasis on cognitive development in the disciplines; (2) the continuity of education over the span of life; and (3) the adaptation of instruction to individual requirements. These influences dictate a form to which evaluative techniques must adapt. The specification of learning outcomes must be well defined in order to evaluate progress toward these goals. For long term projection, a diagnosis of a student's initial state is required. A key task is to determine measures of instructional alternatives to prescribe the most effective sequence of courses. Continuous assessment is necessary to aid in moving to higher and alternative levels. The interaction between individual differences and instructional practices must be known and measured. And finally, the instructional system must be capable of accumulating knowledge from which it can improve its own functioning and come closer to its expressed goals. (LN)

EA

EDO 32647

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR-5-0253
Reprint 46
PA-24
OE-BR

EVALUATION OF INSTRUCTION AND CHANGING EDUCATIONAL MODELS

Robert Glaser
Learning Research and Development Center
University of Pittsburgh

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation of Instructional Programs

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSEIP Occasional Report No. 13, September 1968
University of California, Los Angeles

EA 002 553

EVALUATION OF INSTRUCTION AND CHANGING EDUCATIONAL MODELS[1]

Robert Glaser

Social institutions, whether educational, medical, religious, economic, or political must constantly prove their effectiveness to insure society's support. Acceptable proof of an institution's effectiveness depends largely upon the public attitude toward that institution, an attitude based both upon a respect for authority and tradition and a desire for demonstrated objective proof. (Suchman, 1967). To some extent, the field of educational measurement and evaluation has developed in response to the requirement for objective proof of the effectiveness of the educational enterprise. Furthermore, the demand for evaluation is related to the growing alliance between educational practice and behavioral science and to the pressures which arise from the necessity to make competing social investments. These increasing pressures upon educators, in all parts of the field, to evaluate their activities are one aspect of a growing maturity of the profession and of the commitment of modern society to the belief that its educational problems can be met most effectively through development planned in conjunction with advancing knowledge. However, the

---

[1]Preparation of this chapter was supported under a contract with the Personnel and Training Branch of the Office of Naval Research.

main point I wish to make is that the form which evaluation procedures take is influenced by changes and advances in a given field.

It is reasonable for evaluation practices and procedures to change as the nature of education changes. This is not to imply that educational innovation can completely ignore current standards and procedures of evaluation--a concept that could lead to chaos--but change in educational practice should influence the need for evaluation and the form it takes. Suchman (1967) has pointed out that in the field of public health, evaluation techniques require change as the nature of disease changes. His discussion is pertinent to the theme of this paper. In recent years, acute communicable diseases have been displaced as major causes of death and disability by chronic degenerative diseases. The new diseases are not amenable to the traditional proven methods of environmental sanitation and immunization. The degenerative disease programs, unlike communicable disease programs, cannot depend on either legislative fiat or mass immunization drives but require a greater degree of voluntary public cooperation and long-term programs of prevention and treatment. Evaluation of the control of the new major diseases requires new objectives and the development of new criteria of effectiveness. A heart disease control program, for example, in contrast to a smallpox or diphtheria control program, cannot be evaluated solely in terms of decreasing mortality. Early detection and treatment becomes a new objective, replacing prevention; accomplishment is evaluated and measured in terms of such immediate goals as case finding and the continuity of medical care.

The objectives and evaluation practices of a field are influenced not only by changes in the nature of the field itself but also by changes in the organization and operation of the field. For example, in public health, there is a trend toward broader responsibility for community health; and the dividing line between prevention and treatment is less distinct. Earlier public health services which concentrated on the poor and medically indigent now begin to encompass much larger segments of society. This broad emphasis enlarges the scope of a program's planning, implementation, and evaluation.

As the nature and organization of the field change, so do the attitudes and behaviors of the public, who are both targets of the social enterprise and ultimate determiners of its support. In the early days of the public health movement, the need for environmental sanitation and compulsory immunization did not require proof because the threats from disastrous epidemics were obvious. The feedback and consequences were relatively immediate. Today, the delayed effects of smoking or diet are much less immediate, and evaluation procedures require greater information and proof of the effectiveness of their measures. Today, motivation is a key problem in public health, and one of the primary conditions of motivation is the individual's belief in the effectiveness of the action he is being asked to undertake.

The field of public health provides an apt analogy to the situation which seems to be coming about in educational practice. Consider the three aspects mentioned above: the nature of the field, its organization, and expectations from its user and target

groups. Several forces are changing the nature of educational practice, and of these I shall mention three. One is the increased focus on the cultivation of skill, understanding, and intellectual power in the basic disciplines. Witness the introduction of the massive subject-matter, scholar-based curriculum programs in physics, mathematics, English, history, etc. A second force is the growing conception that education does not have a fixed beginning- or end-point with neat packages of elementary, secondary, and higher education. The stress is less upon third-grade arithmetic or freshman English and more upon the continuity from grade to grade and from age to age and upon a commitment to a transmission of the ability to teach people to teach themselves. The third force is that as we learn more about the psychological and technological foundations of education, individualization of instruction is being viewed less as an ideal and more as a practical enterprise.

Concurrent with the change in the nature of educational activities is the change in the structure, organization, and functioning of these activities and the agencies involved. The trend is toward larger schools, more pervasive educational philosophies, and the integration of social classes in one educational environment. This larger organization and integration deemphasizes local norms and introduces more widely accepted standards of accomplishment and competence. Coupled with this is the necessity for taking account of the increasing heterogeneity of a school by adapting to individual requirements. Another organizational factor that profoundly changes the nature of educational practice is the

continued development of the educational profession and the accruing knowledge in the behavioral sciences.

There is a growing similarity between the public health field and education. Whereas the older diseases had immediately contingent effects that shaped the behavior of the public, the consequences of the newer diseases are more delayed. Perhaps the educational field generally produces effects which have not had immediate consequences mandating immediate action. In this regard, evaluation procedures might provide more immediate feedback of educational outcomes.

A General Instructional Model

Since the nature of educational practice and its organization influences evaluation procedures, it is necessary to present a model of educational practice which can be assumed to underlie any general discussion of the evaluation of instruction. The model I shall describe is one that I believe is likely to come about as a result of the trends I have indicated--the emphasis on cognitive development in the disciplines, the continuity of education over the span of life, the ability to know how to learn and to teach oneself, and the adaptation of instruction to individual requirements. The accomplishment of these objectives suggests an instructional model with the following properties presented as a sequence of operations:

1. Outcomes of learning are specified in terms of the behavioral manifestations of competence and the conditions under which it is to be exercised. This is the platitudinous assertion of the fundamental necessity of describing the foreseeable outcomes of

instruction in terms of certain measurable products and assessable student performance.

2. Detailed diagnosis is made of the initial state of a learner coming into a particular instructional situation. This careful workup of student performance characteristics relevant to the instruction at hand is necessary to pursue further education. Without the assessment of initial learner characteristics, carrying out an educational procedure is a presumption. It is like prescribing medication for an illness without first describing the symptoms. In the early states of a particular educational period, instructional procedures will adapt to the findings of the initial assessment, generally reflecting the accumulated performance capabilities resulting from the long-term behavior history and activity of the learner. The history that is specifically measured is relevant to the next immediate educational step that is to be taken.

3. This immediate instructional step consists of educational alternatives adaptive to the classifications resulting from the initial student educational profiles. These alternative instructional procedures will be selectively assigned to the student or made available to him for his selction.

4. As the student learns, his performance will be monitored and continuously assessed at longer or shorter intervals appropriate to what is being taught. In early skill learning, assessment is quite continuous. Later on, as competence grows, problems grow larger; as the student becomes increasingly self-sustaining, assessment occurs more infrequently. This monitoring serves several purposes: providing a basis for knowledge of results and

appropriate reinforcement contingencies to the learner and a basis for adaptation to learner demands. This learning history accumulated in the course of instruction is called "short-term history" and, in addition to information from the long-term history, provides information for assignment of the next instructional unit. The short-term history also provides information about the effectiveness of the instructional material itself.

5. Instruction and learning proceed in a servomechanismlike, cybernetic fashion, tracking the performance and selections of the student. Assessment and performance are interlinked, one determining the nature and requirement for the other. Instruction proceeds as a function of the relationship between measures of student performance, available instructional alternatives, and learning criteria which are chosen to be optimized. The question of which criteria are to be optimized becomes critical. Is it retention, transfer, the magnitude of difference between pre- and posttest scores, motivation to continue learning including the ability to do so with minimal instructional guidance, or is it all of these? If tracking of the instructional process permits instruction to become precise enough, then a good job can be done to optimize some gains and minimize others unless the presence of the latter gains is desired, expressed, and assessed. The outcomes of learning measured at any point in instruction are referenced to and evaluated in terms of competence criteria and the values to be optimized; provision is always made for the ability of humans to surpass expectations.

6. Inherent in the system's design is its capability for improving itself. Perhaps a major defect in the implementation of educational innovations, especially in the area of individualization, has been the lack of the cumulative attainment of knowledge--on the basis of which the next innovation is better than the one that preceded it.

Given that the changing trends in education will lead to an instructional model somewhat like that just described, the main question to which this paper is addressed is "What are the implications for the nature of evaluation procedures?". I shall examine this question by some elaboration of each of the points just listed.

The Specification of Learning Outcomes

In a system designed to maximize the attainment of certain objectives, the specification of learning outcomes in terms of observable student performance determines how the instructional components are used. Vague specification of desired outcomes leaves little concrete information for the evaluator about what to look for and what to help the system strive to attain. However, interaction between specification of outcomes and instructional procedure provides the basis for redefining objectives. The need for constant revision of objectives is as inherent in the system as is the initial need for defining them. There is a sustained process of clarifying goals, working toward them, evaluating progress, reexamining the objectives, modifying instructional procedures, and clarifying the objectives in the light of evaluated experience. This process should indicate the inadequacies and omissions in a curriculum. The fear of many educators that detailed

specification of objectives limits them to simple behaviors only--those which can be forced into measurable and observable terms--is an incorrect notion if one thinks of them as amendable approximations to our ideals. If complex reasoning and open-endedness are desirable aspects of human behavior, then they need to be recognized and assessable goals. Overly general objectives may force us to settle for what can be easily expressed and measured.

A helpful distinction can be made between the evaluation of procedure and the evaluation of accomplishment. It is possible to evaluate a procedure, such as a difficult surgical operation, and to show that it is being done properly; it is another matter to evaluate its beneficial result. Evaluation of technique may be meaningless without evaluation of its effect, although it is often necessary to show that a new procedure in educational research in the schools is indeed being carried out appropriately. When one neglects the evaluation of technique and moves directly to the evaluation of accomplishment, the effective implementation of the procedure is assumed. One moves from procedural objectives to accomplishment objectives at many points in an instructional sequence. Attaining a procedural objective represents progress toward the accomplishment objective. Even though the two interact and accomplishment objectives are initially established, evaluation designed for the development of an operating instructional system should work from the evaluation of technique to the evaluation of accomplishment objectives--not the other way around as often seems to be the case. In succinct terms, it is necessary

to make sure that the independent variable is in effect before measuring the dependent variable. Of course, in developmental or formative evaluation, assessment of each may suggest changes in the other.

A final point with respect to the specification of objectives relates to the distinction between criterion-referenced and norm-referenced measurement. The measurement of learning outcomes involves the assessment of criterion behavior; implicit in this process is the determination of the characteristics of student performance with respect to specified standards. It can be assumed that regardless of the way a subject matter is structured, some existing hierarchy of sub-objectives indicates that certain performances must be attained as a basis for learning subsequent performance. An individual's competence level falls at some point on this hierarchy of increasing subject-matter competence. The degree to which the individual's measured performance resembles the desired performance at any specified competence level is assessed by referencing his performance to the criterion by some criterion-referenced measure. Criterion levels can be established at any point in instruction where it is necessary to obtain information concerning the adequacy of the learner's performance. The specific behaviors identified at each level of proficiency describe the tasks a student is capable of performing when he achieves this level of knowledge. Performance measured in this way provides explicit information concerning what the individual can and cannot do. Such criterion-referenced measures indicate the content of his behavior and the correspondence between his performance and

the continuum of educational objectives. Measures which assess learner performance in terms of such criterion-referenced standards thus provide information about the competence of a student, independently of reference to the performance of others. In contrast to this procedure, as has been pointed out by Glaser (1963), the general practice in education is to measure achievement by norm referencing rather than by criterion referencing. Norm-referenced measures evaluate the learner's performance in terms of a comparison with the performance of others. Such measures need provide little or no information about the degree of competence exhibited by tested behaviors; they tell that one student is more or less proficient than the other but do not tell how proficient either of them is with respect to the desired learning outcomes. Evaluation in terms of criterion-referenced measures requires that we specify at least minimum levels of behavioral performance that the student is expected to attain or that he needs to attain in order to go on to the next step in an instructional sequence.

Diagnosis of Initial State (Entering Behavior)

The second item in the description of the model refers to the measurement and diagnosis of the initial state or entering behavior with which the learner comes into an instructional situation. Here we appear to be entering the domain of much of the work in the general field of psychological testing and evaluation. It seems obvious, however, that in order to follow through with the model I describe, we must go in the direction pointed to by Cronbach (1957) and by Cronbach and Gleser (1965), that is, to depart from the standard practices of test theory based upon the

basic data of correlations between tests and static criterion variables, and to move toward decision-making procedures based upon the relationships between entering behavior and instructionally manipulated variables. The ultimate purpose of testing in this context is to arrive at decisions with respect to assignment to the instructional treatments defined by these instructional variables.

Evaluation of initial entering behavior involves measuring the products of the long-term history of the learner, which includes what we generally have called aptitudes. These aptitudes have attained importance as fundamental characteristics in the measurement of human behavior because they are useful in predicting long-range criteria such as school and college success. However, the model I describe demands that an additional task for measures of initial behavior be the prediction of very immediate success, that is, success in immediate learning. It can be postulatd that if the criteria for aptitude test validation had been immediate learning success rather than some long-range criteria, the nature of today's generally accepted aptitude batteries would be quite different. This postulation seems likely since factorial studies of the changing composition of abilities over the course of learning (Fleishman, 1965) show different abilities involved at the beginning and end of the course of learning. Thus, while it is useful to forecast over the long range, our instructional model also requires measures which are closely related to more immediate learning criteria, that is, success in initial instructional steps. Current types of measured aptitude may be limited in that they are operationally designed to predict over the long period, given reasonably nonadaptive forms of educational treatment.

Aptitude tests or general psychometric reference tests resulting from factor analyses of aptitude tests would not be expected to correlate very highly with individual differences in learning and thereby would not be useful for the placement of individuals in alternate instructional treatments. As Jensen (1967) has pointed out, the predictive power of tests like the Primary Mental Abilities test is due to the fact that they sample learned behavior and therefore reflect something about the rate of learning in a given environment. They also measure the acquisition of broad verbal or symbolic capabilities (mediational systems), which play an important role in enabling an individual to generalize and solve problems. However, such standard psychometrically developed tests, as a result of the way in which they have been validated and evaluated, are more closely related to the products of learning which they predict, such as ability in school subjects, than they are to the kinds of variables generally dealt with in the learning laboratory; conceivably they are relevant to instructional manipulation and educational alternatives. Evidence for this lack of utility of general psychometric measures with respect to instructional decisions comes from the line of studies dealing with correlations between psychometric variables and learning measures which was begun in 1946 by Woodrow's classic article. Woodrow showed data from laboratory and classroom experiments which indicated that the correlations between intelligence measures and ability to learn, in the sense of ability to improve with practice, were generally insignificant and often closet to zero. More recently, this work has been followed up by Gulliksen and his students, for

example, Stake (1961) and Duncanson (1964); but the results obtained are not clear-cut, and Woodrow's basic point has not been clearly disclaimed.

It seems that approximately five categories of entering behavior would require measurement for instructional decision-making (Travers, 1963): (a) the extent to which the individual has already learned the behavior to be acquired in instruction, i.e., previously attained achievement in the skills and knowledge to be taught, (b) the extent to which the individual possesses the prerequisites for learning the behavior to be acquired, for example, knowing how to add before learning to multiply, (c) learning set variables, which consist of acquired ways of learning which facilitate or interfere with new learning procedures under certain instructional conditions, for example, prior success in being impulsive versus being reflective, (d) specific ability to make discriminations necessary in subsequent instruction, for example, musical aptitude or spatial visualization, and (e) general mediating abilities as measured by general tests of verbal or symbolic intelligence.

Instructional Alternatives

From the initial measurement, instructional alternatives are available to the student. But what are these instructional alternatives, where do they come from, and how are they developed? In other words, on what basis do different instructional treatments differ so as to be adaptive to individual requirements? This is a significant problem fundamental to psychologically-based instructional design but which, in this paper emphasizing evaluation, can only be mentioned. Some goals seem easy to achieve, such as adapting

to the student's present level of accomplishment, his mastery of prerequisites, the speed at which he learns including the amount of practice he requires, and his ability to learn independent of highly structured situations. Adaptation to treatments differing in these respects, which are shown to be related to measured aspects of entering behavior, might be able to provide a significant beginning for effective adaptation to individual differences. However, in designing instructional alternatives, it is difficult to know how to use other variables which come out of learning theory (such as requirements for reinforcement, distribution of practice, the use of mediation and coding mechanisms, and stimulus and modality variables, e.g., verbal, spatial, auditory, and visual presentation; and more needs to be known about their interaction with individual differences.

If one assumes that measures of entering behavior and instructional treatments are both available, then at our present state of knowledge, empirical work must take place to determine those measures most efficient for assigning individuals to treatment classes. The task is to determine those measures that have the highest discriminating potential for allocating between treatments and then determine their intercorrelations so that they can be combined in some way and all of them need not be used. This task seems to be a reasonably typical multivariate problem. As a result of the initial diagnostic or placement decision, the universe or sample of students involved is reduced to subsets, allocable to the various available instructional treatments. These initial decisions will be corrected by further assignments as learning proceeds

so that the allocation procedure becomes a multistage decision process which defines an individualized instructional path.

Continuous Assessment

The next item in the model indicates that as a student proceeds to learn his performance will be monitored, and at appropriate intervals, measures of this performance will be summarized and indexed. In contrast to the long-term history used for initial placement, the measures obtained in learning are called the short-term history, even though prolonged use of the model may fuse the two items to some extent. Here again, the problem of what instructional alternatives are made available is of major concern. Of equal importance are the kinds of measures to be obtained in the course of learning.

The kinds of measures of learning progress one usually obtains, and on which instructional decisions are made, consist of test score information which measures the frequency of correct responses, errors in relation to some performance standard, and the speed of performance. Less frequently, measures of transfer and generalization are specifically developed. Perhaps, to some extent, this is done when one selects a set of test items which are derived from the same universe of subject-matter content but are not the same sample as was used in initial learning. Of special interest in the assessment of short-term history are measures that are being suggested by experimental work on learning; these are measures which can be obtained in the course of learning and may be predictive of future learning requirements. Two examples may give the

flavor of this. One comes from the work of Zeaman and House (1967) on a theory of discrimination learning accounting for the performance of retarded children learning to solve two-choice visual discrimination problems, such as may be involved in letter or numeral discrimination. The theory postulates a chain of two responses for problem solution: the first, paying attention to the relevant stimulus dimensions, and the second, the correct selection of the positive cue of the relevant dimensin. They ask whether individual differences in empirical learning curves are attributable to differences in the speed of acquisition or to some underlying process such as attention. The data they obtain show wide individual differences in learning curves, with higher IQ subjects doing better than the lower; however, the important differences in the curves between the brighter and duller subjects is not the sloped of the curve, i.e., the rate of learning, but the length of the initial plateau. Thus, it is not the rate of improvement, once it starts, that distinguishes bright and dull but how long it takes for improvement to begin. The length of time for improvement to begin is considered an attentional variable and suggests, at least with respect to the concerns of this paper, that the measurement of plateau length rather than rate of improvement is a sensitive measure of discrimination learning.

The second example is a study performed in my own laboratory by Wilson Judd (1967) on paired-associate learning. The interest here was on response latency, that is, the interval between the onset of a stimulus and the occurrence of a response, as an index of learning. Hull, in his theory and experimental work, strongly

suggested latency as a measure of habit strength. Our study investigated changes in the latency measure over the course of learning, from initial learning through a criterion of nearly perfect performance, and then through overlearning. Throughout this course, frequency of correct response increased to criterion and then continued at asymptote through overlearning. In contrast, latency showed no change and remained constant as correct response probability increased from chance to near 1.0; however, during the overlearning period, while response probability remained constant, latency showed a significant and sustained decrease, presumably related to the consolidation of learning during the overlearning period. The suggestion from this work is that the latency measure, as a short-term learning history variable, seems to detect aspects of learning not detectable from response frequency and may be related to and predictive of future retention. With the talk about the possibility of computer-assisted instruction, latency measures would be easy to obtain and be available for instructional decision-making.

The work of Jensen (1967) on individual differences in learning variables is also relevant here. His factor analyses of learning tasks of the kind used in the learning laboratory showed interesting results. For example, two types of learning which on the surface look very much alike, serial learning and paired-associate learning, were not found to be significantly intercorrelated, even when the stimulus materials were the same in both tasks. In addition these was little transfer between the two tasks. On the other hand, serial learning was found to have much in common with memory

span. Jensen also found that in serial learning, individual differences in original learning are not highly correlated with individual differences in subsequent learning. The reliability of measures of learning variables for individual difference work posed problems for Jensen. In general, the point to be made is that the psychometrics of learning measures poses itself as a new evaluation task.

Adaptation and Optimization

The fifth item in the instructional model indicates that the assessment of behavior during learning and instructional assignment is interlinked in a series of adaptive stages. Two points are appropriate. First, information about learning relevant to this kind of instructional model should come primarily from the interaction effects generally neglected in studies of learning. As Cronbach and Gleser (1965) have pointed out, the learning experimentalist assumes a fixed population and hunts for the treatment with the highest average and least variability. The correlational psychologist has, by and large, assumed a fixed treatment and hunted for aptitude which maximizes the slope of the function relating outcome to measured aptitude. The present instructional model assumes that there are strong interactions between individual measurements and treatment variables; and unless one treatment is clearly the best for everyone, as may rarely be the case, then treatments or instructional alternatives should be differentiated in a way to maximize their interaction with performance variables. If this assumption is correct, then individual performance measures that have high interactions with learning variables and their

associated instructional alternatives are of greater importance than measures which do not show these interactions. This forces us to break out the error term in learning experiments so that the subject-by-independent-variable interaction can be evaluated. When this interaction is shown to be negligible, the learning variable can then be used in instruction without correcting its values to individual differences. It seems that the model I have described will require major experimental research to determine the extent to which instructional treatments need to be qualified by individual difference interactions. The search for such interactions has been a major effort in the field of medical diagnosis and treatment and seems to be so in education (Lubin, 1961).

Second, the continuous pattern of assessment and instructional prescription, and assessment and instructional prescription again, can be represented as a multistage decision process where decisions are made sequentially and decisions made early in the process affect decisions made subsequently. The task of instruction is to prescribe the most effective sequence. Problems of this kind in other fields, such as electrical engineering, economics, and operations research, have been tackled by mathematical procedures applied to optimization problems. Essentially, optimization procedures involve a method of making decisions by choosing a quantitative measure of effectiveness and determining the best solution according to this criterion with appropriate constraints. A quantitative model is then developed into which values can be placed to indicate the outcome that is produced when various values are introduced.

An article by Groen and Atkinson (1966) has pointed out that the kind of instructional model I have described is set up for this kind of analysis. There is a multistage process which can be considered as a discrete N-stage process. At any given time, the state of the system, i.e., the learner, can be characterized. This state, which is probably multivariate and described by a state vector, is followed by a decision which also may be multivariate; the state is transformed into the new updated state. The process consists of N successive states where at each of the N-1 stages a decision is made. The last stage, the end of a lesson unit, is a terminal stage where no decision is made other than whether the terminal criteria have been attained. The optimization problem of major concern in this process is finding a decision procedure for deciding which instructional alternatives to present at each stage, given the instructional alternatives available, the set of possible student responses to the previous lesson unit, and specification of the criteria to be optimized for the terminal stage. This decision procedure defines an instructional strategy and is determined by the functional relationship between (a) long- and short-range history and (b) student performance at each state and at the terminal stage.

Groen and Atkinson (1966) point out that one way to find an optimal strategy is to enumerate every path of the decision tree generated by the multistage process. Obviously, this can be improved upon by the use of adequate learning models which can reduce the number of possible paths that can be considered. In order to reduce these paths still further, Bellman, (1957) and

Bellman & Dreyfus, (1962), refer to dynamic programming procedures as useful for discovering optimal strategies and hence for providing a set of techniques for reducing the portion of the tree that must be searched. I am intrigued by this and suggest that it is an interesting approach for evaluation theory to consider, although some initial experimentation has not been overwhelmingly successful and, perhaps, slightly discouraging.

In order to carry out such an approach, we need only to do two trivial things: first, obtain quantitative knowledge of how the system variables interact, and second, obtain agreed upon measures of system effectiveness. Upon the completion of these two simple steps requiring, respectively, knowledge and value judgment, optimization procedures can be carried out. It has been shown that relative to the total effort needed to achieve a rational decision, the optimization procedure itself often requires little work when the first two steps are properly done (Wilde & Beightler, 1967). We are thrown back to the tasks we have always known that we must confront: (a) knowledge and description of the instructional process and (b) the development of evaluation measures.

In the first task the question is what kinds of experimental tactics and learning theory are most useful for discovering individual-difference-learning-variables relationships required to develop an instructional system. Fortunately, there is a growing commitment in learning theory to the individual case--recognized but not incorporated to any extent by Hull, certainly urged upon us by Skinner and associates, and well recognized in the recent

information-processing computer simulation models of human behavior. There seems little doubt that one major test of the adequacy of competing learning theories will be the extent to which they incorporate individual differences.

The second task refers to the fact that in the educational model described, criterion measures and what is to be optimized become critical. If tracking the instructional process permits instruction to become precise enough, a good job can be done to maximize some gains and minimize others but some criteria may be minimized inadvertently unless the presence of the latter are desired, expressed, and assessed. In this regard, it seems almost inescapable that we abandon only norm-referenced measurement and develop more fully criterion-referenced measures, measures which assess performance on a continuum of competence and growth in the area under consideration. In addition, serious attempts must be made to measure what has been heretofore so difficult; such aspects as transfer of knowledge to new situations, problem solving, and self-direction--those aspects of learning and knowledge that are basic to an individual's capability for continuous growth and development.

Evolutionary Operation

The final item in my model refers to the capability of an instructional system to gather information and accumulate knowledge from which it can improve its own functioning and come closer to its expressed goals. I think the current notion of "formative" evaluation inherent in programmed instruction and presently being discussed more generally in curriculum evaluation is

a major step along these lines (Cronbach, 1963). The industrial concept of "evolutionary operation" is relevant here (Box, 1957). The underlying rationale of this concept states it is seldom efficient to run an industrial process to produce a product alone; the process should produce the product plus information about how to improve it.

In closing the remarks in this paper, I can think of nothing better than to quote the end of Cronbach's 1963 article entitled "Evaluation for Course Improvement." He writes:

> Old habits of thought and long-established techniques are poor guides to the evaluation required for course improvement. Traditionally, educational measurement has been chiefly concerned with producing fair and precise scores for comparing individuals. Educational experimentation has been concerned with comparing score averages of competing courses. But course evaluation calls for description of outcomes. This description should be made on the broadest possible scale, even at the sacrifice of superficial fairness and precision.
>
> Course evaluation should ascertain what changes a course produces and should identify aspects of the course that need revision.
>
> . . . Evaluation is a fundamental part of curriculum development, not an appendage. Its job is to collect facts the course developer can and will use to do a better job, and facts from which a deeper understanding of the educational process will emerge.

Conclusion

I have stated the thesis that changing educational practices require changes in our theories and techniques of evaluation. In a general model of an emerging instructional process, I have itemized six educational practices and suggested the considerations for evaluation and measurement which each raises. They are the following:

1. With respect to the specification of learning outcomes, the following are required: (a) behavioral definition of goals, evaluating progress toward these goals, and clarifying these goals in the light of evaluated experience, (b) prior evaluation of educational procedures, insuring they are in effect before assessing educational accomplishment, and (c) development of techniques for criterion-referenced measurement.

2. For the diagnosis of initial state, what is required is determination of long-term individual differences that are related to adaptive educational alternatives.

3. For the design of instructional alternatives, a key task is to determine measures which have the highest discriminating potential for allocating between instructional treatments.

4. For continuous assessment, discovery of measurements of ongoing learning which facilitate prediction of the next instructional step is required.

5. For adaptation and optimization, the instructional model requires: (a) the detailed analysis of individual-difference by instructional-treatment interactions and (b) the development of procedures like the optimizing methods so far used in fields other than education.

6. For evolutionary operation, we require a systematic theory or model of instruction into which accumulated knowledge can be placed and then empirically tested and improved.

REFERENCES

Bellman, R. Dynamic programming. Princeton: Princeton University Press, 1957.

Bellman, R., & Dreyfus, S. E. Applied dynamic programming. Princeton: Princeton University Press, 1962.

Box, G. E. P. Evolutionary operation: A method for increasing industrial productivity. Applied Statistics, 1957, 6, 81-101.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J. Evaluation for course improvement. Teachers College Record, 1963, 64, 672-683.

Cronbach, L. J., & Gleser, G. C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1965.

Duncanson, J. P. Intelligence and the ability to learn. Princeton: Educational Testing Service, 1964.

Fleishman, E. A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), Training research and education. New York: Wiley and Sons, 1965, 137-176.

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.

Groen, G. J., & Atkinson, R. C. Models for optimizing the learning process. Psychological Bulletin, 1966, 66, 309-320.

Jensen, A. R. Varieties of individual differences in learning. In R. Gagne (Ed.), Learning and individual differences. Columbus, Ohio: Charles Merrill Books, 1967, 117-135.

Judd, W. A. The effects of task characteristics on response latency during paired-associate learning. Unpublished doctoral dissertation, University of Pittsburgh, 1967.

Lubin, A. The interpretation of significant interaction. Educational and Psychological Measurement, 1961, 21, 807-817.

Stake, R. E. Learning parameters, aptitudes, and achievements. Psychometric Monographs, 1961, No. 9.

Suchman, E. A. Principles and practice of evaluative research. In J. T. Doby (Ed.), An introduction to social research. New York: Appleton-Century-Crofts, 1967, 327-351.

Travers, R. M. W. Essentials of learning: An overview for students of education. New York: Macmillan, 1963.

Wilde, D. J., & Breightler, C. S. Foundations of optimization. Englewood Cliffs, N. J.: Prentice-Hall, 1967.

Woodrow, H. A. The ability to learn. Psychological Review, 1946, 53, 147-158.

Zeaman, D., & House, B. J. The relation of IQ and learning. In R. Gagne (Ed.), Learning and individual differences. Columbus, Ohio: Charles Merrill Books, 1967, 192-212.